Depth-Based Whole Body Photoplethysmography in Remote Pulmonary Function Testing

Vahid Soleimani, *Student Member, IEEE*, Majid Mirmehdi, *Senior Member, IEEE*, Dima Damen, *Member, IEEE*, Massimo Camplani, *Member, IEEE*, Sion Hannuna, Charles Sharp, and James Dodd

Abstract*—*Objective: **We propose a novel depth-based photoplethysmography (dPPG) approach to reduce motion artifacts in respiratory volume–time data and improve the accuracy of remote pulmonary function testing (PFT) measures. *Method:* Following spatial and temporal calibration of two opposing RGB-D sensors, a dynamic three-dimensional model of the subject performing PFT is reconstructed and used to decouple trunk movements from respiratory motions. Depth-based volume–time data is then retrieved, calibrated, and used to compute 11 clinical PFT measures for forced vital capacity and slow vital capacity spirometry tests. *Results:* A dataset of 35 subjects (298 sequences) was collected and used to evaluate the proposed dPPG method by comparing depth-based PFT measures to the measures provided by a spirometer. Other comparative experiments between the dPPG and the single Kinect approach, such as Bland–Altman analysis, similarity measures performance, intra-subject error analysis, and statistical analysis of *tidal volume* and *main effort* scaling factors, all show the superior accuracy of the dPPG approach. *Conclusion*: We introduce a depth-based whole body photoplethysmography approach, which reduces motion artifacts in depth-based volume–time data and highly improves the accuracy of depth-based computed measures. *Significance:* The proposed dPPG method remarkably drops the L_2 error mean and standard deviation of $FEF_{50\%}$, $FEF_{75\%}$, $FEF_{25-75\%}$, *IC*, and *ERV* measures by half, compared to the single Kinect approach. These significant improvements establish the potential for unconstrained remote respiratory monitoring and diagnosis.**

***Index Terms*—3-D body reconstruction, motion artifacts reduction, motion decoupling, depth-based photoplethysmography (dPPG), forced vital capacity (FVC), lung function assessment, pulmonary function testing, slow vital capacity (SVC), spirometry.**

Manuscript received September 4, 2017; revised October 19, 2017; accepted November 17, 2017. Date of publication December 11, 2017; date of current version May 18, 2018. This work was supported by the University of Bristol Alumni Foundation. (*Corresponding author: Vahid Soleimani.)*

V. Soleimani is with the Department of Computer Science, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: vahid.soleimani@bristol.ac.uk).

M. Mirmehdi, D. Damen, M. Camplani, and S. Hannuna are with the Department of Computer Science, University of Bristol.

C. Sharp and J. Dodd are with the Academic Respiratory Unit, Southmead Hospital.

Digital Object Identifier 10.1109/TBME.2017.2778157

I. Introduction

LUNG function diseases, e.g., Chronic Obstructive Pulmonary Disease (COPD), Asthma and lung fibrosis, affect many people and are major causes of death worldwide [1]. Spirometry [2] and whole body plethysmography [3] are traditional and clinically approved methods for pulmonary function testing (PFT), but spirometry is more prevalent and broadly used in clinical environments due to its relative affordability, portability and accuracy.

Forced vital capacity (FVC) and slow vital capacity (SVC) are two primary clinical protocols undertaken with a spirometer that vary in the pattern of breathing into the spirometer. FVC is comprised of a maximal inhalation followed by a forced maximal exhalation, and SVC a maximal inhalation followed by a slow, controlled, maximal exhalation. Both tests start with a few cycles of normal breathing, called *tidal volume*, followed by the intended lung function test, called *main effort*. Various clinical PFT measures are estimated within FVC and SVC protocols [2], [4]. These measures, i.e., *FVC*, *FEV1*, *PEF*, ..., $FEF_{25-75\%}$ (FVC measures) and *VC*, *IC*, *TV*, *ERV* (SVC measures), and their combinations, e.g., *FEV1*/*FVC*, are used for the diagnosis of *obstructive* and *restrictive* lung diseases. Airway resistance, defined as lung pressure divided by airflow, is another measure which can be used in the diagnosis of other pulmonary diseases, such as Respiratory Syncytial Virus. However, this study only focuses on the estimation of PFT measures, which can be directly validated by measures provided by a spirometer.

Despite its reliability and accuracy, spirometry has certain drawbacks, such as being intrusive and difficult to deal with for all subjects, particularly for children and the elderly. Since it requires the patient's cooperation during the test, cognitively impaired people may find it troublesome to coordinate with it. Spirometry is a rather expensive approach given the price of pneumotach and the required disposable accessories (mouthpiece and nose clip), and it also requires specialist training. Further, a pneumotach must be calibrated before each session to be able to measure accurately. Thus, remote respiratory sensing has recently become very popular and numerous approaches have been proposed for lung function assessment [5], [6], respiration resistance [7], [8], and tidal volume respiratory monitoring and breathing rate estimation [9]–[15], based on time-of-flight

[5], [6], [8], [12], [13] and structured light depth sensors [7], [9]–[11] and RGB video cameras [14], [15]. Some of these are briefly considered in Section II.

Among all the existing related studies, we are only aware of Ostadabbas *et al.* [7], [8] and our previous works [5], [6], which performed clinical respiratory assessment. Ostadabbas *et al.* [7], [8] mainly focused on airway resistance estimation and estimated two measures, *FVC* and *FEV1*, in [7]. In [6], we introduced a remote lung function assessment approach to estimate 11 PFT measures using a single depth sensor. The PFT measures were computed from a calibrated depth-based volume–time data, obtained by estimating the variation of chest volume per frame. The calibration process linearly scaled the estimated chest volume to the real lung volume using intra-subject scaling factors learnt in a training phase. In order to compute the scaling factors and PFT measures, several keypoints were automatically detected from the volume–time data. The medical significance of our remote lung function assessment approach has been reported in [16], [17].

Our previous approaches [5], [6] were based on a single depth sensor, which made them very sensitive to the subject's trunk motion during the PFT test. Although subjects were asked to be completely still, most of them inevitably moved their trunk, especially during the deep forced *inhalation–exhalation*. This body movement is a natural reaction of the human respiratory system when required to maximally inhale and exhale. Decoupling the trunk motion and the chest-surface respiratory motion under such circumstances is potentially impossible. Similar body motion artifacts have been also reported in [7], [8], [10] where the main solution was to constrain the body movement, which is neither easy to achieve, nor particularly comfortable for patients.

In this paper, we propose a whole body depth-based photoplethysmography (dPPG) approach for lung function assessment, in which we use two opposing Kinect V2 sensors to decouple trunk movements from respiratory motions by constructing a dynamic full 3-D model of the subject during PFT performance. We validate our proposed method by comparing our PFT measures, computed for 35 healthy subjects (298 sequences), to the measures obtained from a spirometer.

The most significant novelties of this work are that it introduces the concept of motion decoupling into the remote, vision-based respiratory sensing area and demonstrates its efficacy and achievement in pulmonary function testing. Constraining the body's natural reaction to deep forced *inhalation–exhalation* can prevent subjects from performing their best breathing effort and would therefore affect their lung function measures. Unlike all previous remote approaches which restrict the subject's movement during their tests [5]–[17], our proposed method allows subjects to perform PFT as routine spirometry procedures without restricting the subject's natural body reactions at the *inhalation–exhalation* stages. Our contribution to the state-of-the-art is therefore to facilitate remote respiratory monitoring and diagnosis without unduly constraining patients.

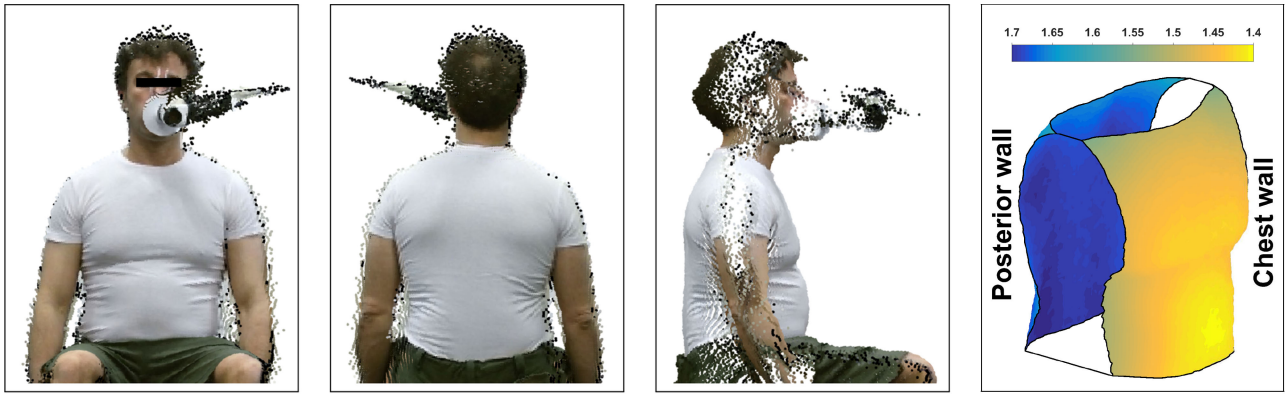
We demonstrate the accomplishments of our dPPG approach by, (*a*) achieving significant improvements in FVC and SVC measures compared to the single Kinect approach, (*b*) improving volume–time data calibration accuracy by computing more accurate similarity measures and reducing intra-subject scaling factor learning error, (*c*) computing more consistent and stable *tidal volume* and *main effort* scaling factors, which increases the depth-based PFT measures reproducibility, and (*d*) achieving higher correlation between *tidal volume* and *main effort* scaling factors confirmed by performing a comparative statistical analysis across 35 subjects.

Fig. 1. The proposed system for performing PFT with 2 opposing Kinects.

Next, Section II briefly reviews the state-of-the-art vision-based respiratory sensing methods, relevant works in reducing motion artifacts in PPG signals, and also multiple Kinect calibration and registration methods. Then, Section III describes the proposed dPPG methodology in which for each frame, the two point clouds from two opposing Kinects (see Fig. 1) are synchronised and registered, and the subject's 3-D trunk model is constructed. A pair of volume–time data sets, automatically estimated from the chest and posterior regions, are then combined to retrieve the final depth-based volume–time data. Several keypoints are then automatically extracted from this volume–time data which are used to compute *tidal volume* and *main effort* calibration scaling factors and PFT measures. Since these scaling factors are subject-specific, we train our proposed system for each subject, which enables our method to compute PFT measures using only depth-based volume–time data afterwards. Experimental results are reported in Section IV. In addition to evaluating the depth-based PFT measures against a spirometer, we statistically analyse intra-subject scaling factors and assess their stability and generalizability for all subjects. The paper is concluded in Section V.

II. Related Works

Vision-based respiratory sensing – Ostadabbas *et al.* [7] estimated airway resistance and computed *FVC* and *FEV1* measures for five healthy subjects using a Kinect. Subjects were asked to blow through various straws to induce varied airway resistance while their lung volume was measured over time. For the PFT measures evaluation, an average 0.88 correlation with the spirometer was reported for *FEV1*. They expanded this study in [8] and used a time-of-flight depth sensor along with a pulse oximeter to determine the severity of airway obstruction as mild, moderate or severe. They reported 76.2% and 80% accuracy in



(a) front view (b) back view (c) side view (d) 3-D trunk model

Fig. 2. (a)–(c) 3-D reconstructed model of a subject performing PFT from different points of view. (d) 3-D reconstructed model of trunk for this subject.

detecting airway obstruction of 14 healthy subjects (with simulated airway obstruction) and 14 patients, respectively. Both studies [7], [8] restricted the trunk movement by asking their subjects to press their back against a wall during the test.

Rihana *et al.* [12] estimated the respiratory signal using distance information of a manually selected ROI on the subject's chest. They evaluated their method on 10 healthy subjects and reported maximums of 85% and 50% correlations against a respiratory belt in normal and high-frequency breathing, respectively. Using a depth sensor, Transue *et al.* [13] reconstructed the chest-wall surface per frame to estimate tidal volume breathing. For each subject, they calibrated the estimated chest volume using a Bayesian network, trained on spirometer and Kinect data. They reported $92.2\% - 94.19\%$ accuracy in tidal volume estimation for 4 healthy subjects. Similarly, Aoki *et al.* [9] and Yu *et al.* [10], computed the subject's chest volume variations in depth sequences to estimate the airflow signal, and respectively reported 0.98 and 0.96 correlation against their groundtruth. Seppanen *et al.* [11] estimated airflow signal using multi-input–single-output models fed by the data acquired using a depth sensor. Their best correlation against a spirometer was $R^2 = 0.93$.

Reyes *et al.* [15] acquired chest breathing motions using a smartphone camera, and estimated tidal volume breathing on a PC using average pixel intensity in R,G, and B channels. A correlation of 0.95 was reported for the estimated tidal volume against a spirometer for 15 healthy subjects.

Motion artifacts reduction in PPG signals – PPG signals, obtained from wearable devices such as pulse oximeters and wrist-bands [18], [19], or by remote approaches [20], [21], are used to extract heartbeat rate, arterial oxygen saturation (SpO_2) and breathing rate. PPG signals can also be corrupted by a subject's movement during the test. Although motion artifacts reduction in regular PPG signals has been widely investigated [22], [23], these signals are quite different in their nature and behaviour compared to the spirometry volume–time data.

The most relevant work to this study, in terms of motion artifacts reduction in respiratory signals is [24], in which Shao *et al.* exploited an HD video camera to estimate breathing frequency using two tiny ROIs ($40\times$ 40 pixel), manually selected from the top of the shoulders. In order to reduce motion artifacts, these two ROIs were tracked using shoulders' gradient information. The ROI size was chosen as a trade-off between tracking and breathing rate estimation accuracy. Although this approach can track up–down shoulder movements and reduce motion artifacts in tidal volume breathing signals, it is not able to track forward–backward trunk motions during deep and forced inhalation–exhalation. Further, spirometry volume–time data cannot be estimated using such small ROIs.

Multiple Kinect calibration and registration – To the best of our knowledge, there are only a few works on calibrating multiple Microsoft Kinect V2 RGB-D sensors, e.g., [25], [26], possibly due to specific hardware and software needs, e.g., an individual PC for each sensor would be required.

To calibrate multiple Kinect V2 sensors to capture a space of about $1.5m \times 1.8m \times 1.5m$, Beck and Froehlich [25] tracked a moving chessboard with a motion capture system to fill a lookup table with 2000 reference samples over 20–30 minutes. This lookup table was then interpolated and used in the reconstruction stage. Kowalski *et al.* [26] presented a 3-D data acquisition and registration system, which calibrates up to four Kinect V2 sensors in a two-step procedure, involving a rough estimation step using their self-designed markers, and a refinement step, using an adapted iterative closest point (ICP) algorithm which requires sufficient overlap between the sensors. Their work demands a cumbersome calibration stage, which requires manual labelling of marker locations. They did not report quantitative results on their spatial and temporal registration accuracy.

III. Proposed Methodology

A. Reconstructing the 3-D Trunk Model

In order to compute depth-based PFT measures correctly, especially the timed measures such as *FEV1*, it is necessary for the depth-based volume–time data to have a constant and high sampling rate. Since it is impossible to trigger multiple Kinects simultaneously, an exact frame level synchronisation between them cannot be achieved. Thus, the more Kinects that are used, the greater the temporal synchronization error would be. Our

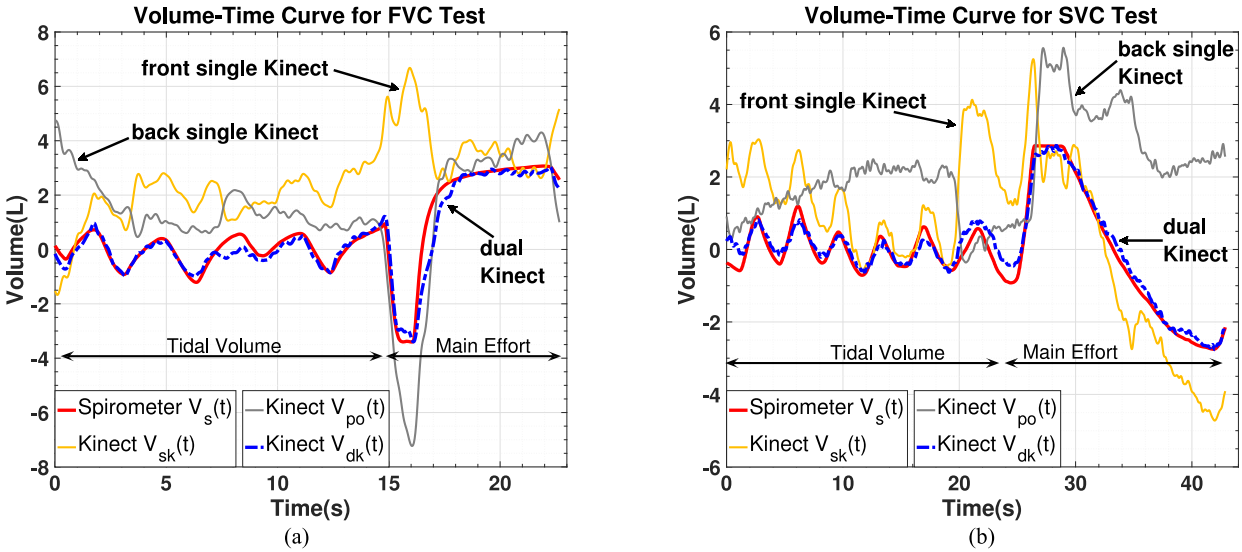


Fig. 3. Comparing FVC (a) and SVC (b) volume–time curves of dual Kinect $V_{dk}(t)$, front Kinect $V_{sk}(t)$, and back Kinect $V_{po}(t)$ to spirometer $V_s(t)$.

dual-Kinect 3-D data acquisition and registration system [27] reconstructs an almost complete 3-D model of a subject, performing the routine PFT in a sitting position, at full frame rate (*30fps*). Deploying only two Kinect sensors, (*a*) minimises the temporal and spatial alignment error, (*b*) reduces the system setup and calibration effort, (*c*) keeps system costs low, and (*d*) minimises the overall operation space. With this topology, there would be no overlapped views of the scene and the thoracoabdominal regions occluded by the arms are not considered pertinent to volume estimation accuracy. Fig. 2(a)–(c) shows the 3-D reconstructed model of a subject's upper-body from different viewpoints performing PFT.

Temporal synchronisation – Intra-Kinect synchronisation, to temporally align corresponding RGB, depth and skeleton data from each Kinect running on a different PC, is implemented by synchronising the system time of the two locally networked PCs using Network Time Protocol (NTP).

Registration – As there is no overlap between the point clouds of two facing Kinects, ICP-based calibration approaches to align the point clouds cannot be employed. Thus, we apply an automatic, fast and accurate optical calibration method, in which three double-sided chessboards are placed at different depths from the Kinects (to improve the spatial registration accuracy) to estimate the rigid transformation parameters, i.e., translation and rotation matrices. These parameters are then used in the reconstruction stage to register the two Kinects' point clouds to a joint coordinate system at frame-level. Note, as long as the position of the depth sensors remains fixed, these calibration parameters remain valid. The registration accuracy of the proposed method was quantitatively assessed by measuring geometrical specification of three boxes of known size. The average error range across 3 boxes at 3 different placements was 0.21 – 0.84 cm. The synchronisation and registration methodology, and registration accuracy, is comprehensively reported in [27] and the source code is publicly available.[1]

[1]https://github.com/BristolVisualPFT/

Fig. 2(d) presents the final 3-D reconstructed model of the subject's trunk after removing head and limbs using a 3-D mask, automatically computed from body skeletal data.

B. Volume–Time Data Retrieval

After registering the models of the chest and posterior walls to a joint real-world coordinate system for each frame of the sequence, a pair of volume–time curves, i.e., $V_{ch}(t)$ and $V_{po}(t)$, are computed using an averaging-based method. As an en bloc object, the subject's trunk movements are reflected on both the chest and the posterior walls, whereas the breathing motions mainly appear on the chest wall, with the posterior wall considerably less affected. Taking this into consideration, the trunk movements can be cancelled out by subtracting the motions of the chest and the posterior walls per frame, due to their similarity in direction and magnitude. However, this subtraction intensifies the breathing motions because expansion and contraction of the lungs move the chest and the posterior walls in nearly opposite directions. Thus, the final depth-based volume–time curve, i.e., $V_{dk}(t)$, is computed as $V_{dk}(t) = [V_{po}(t) - V_{ch}(t)]$. To compare our proposed method with the single Kinect approach, the single Kinect volume–time curve is defined as $V_{sk}(t) = V_{ch}(t)$. Note that, $V_{po}(t)$ does not present any meaningful or useful information on its own.

We improve the data filtering method in three ways compared to [6]. First, we chose not to apply a Bilateral smoothing filter, as we noticed it eliminates subtle respiratory motions and affects the final PFT measures, especially the flow-based measures, i.e., *PEF* and $FEF_{25\%}$, $FEF_{50\%}$, $FEF_{75\%}$ and $FEF_{25-75\%}$. Second, we realised that applying a moving-averaging filter [6] over-smooths the *main effort* part of the volume–time curve and increases the error in flow-based PFT measures. Thus, in this work we use a 4th order Butterworth low-pass filter to smooth $V_{dk}(t)$ and $V_{sk}(t)$, similar to [8]. Third, we perform a twofold volume–time curve filtering with two different cut-off frequencies. In the first stage, in order to (*a*) identify the keypoints accurately, (*b*) align Kinect and spirometer volume–time curves

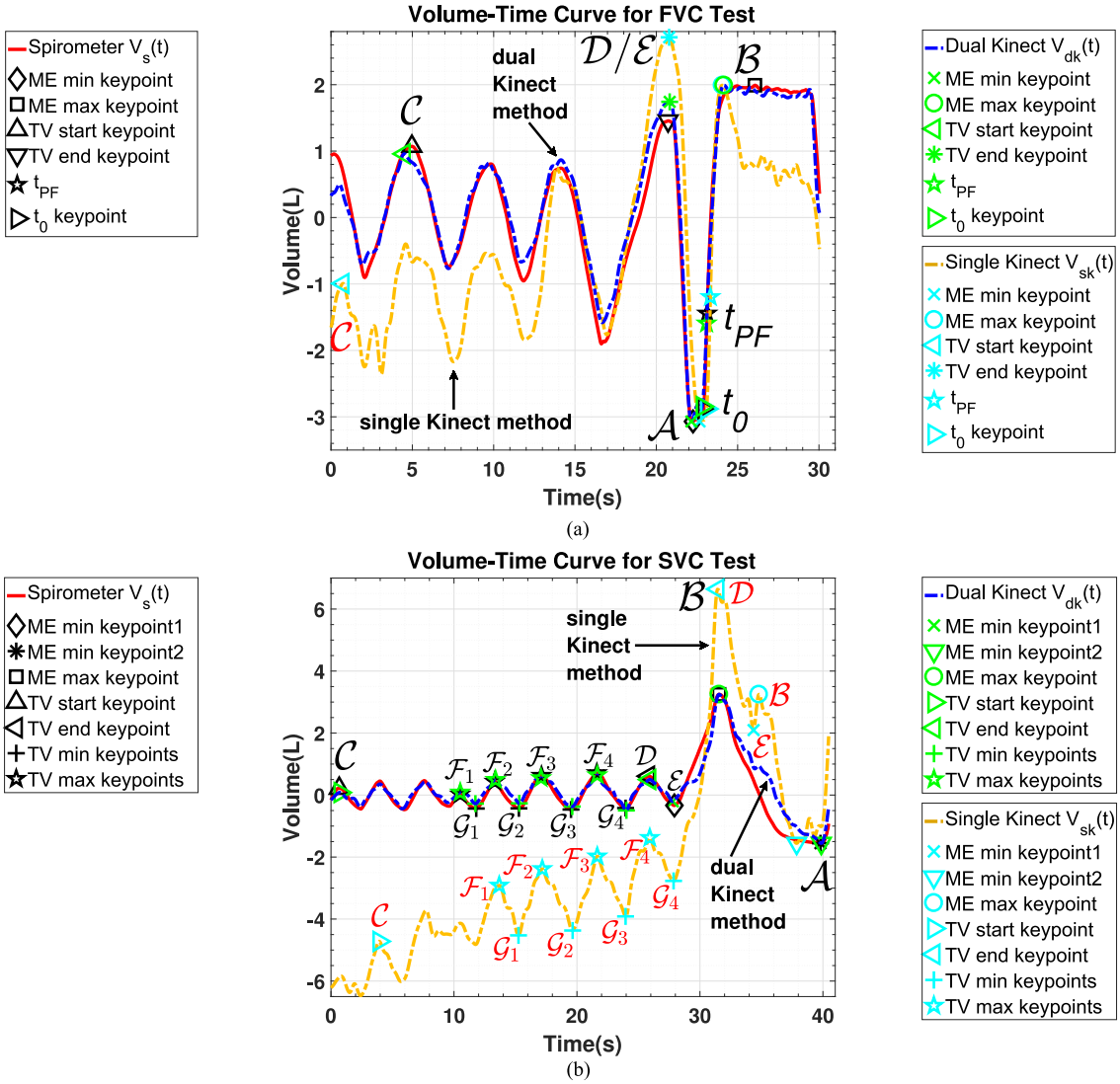


Fig. 4. Comparison of volume–time curves of dual Kinect $V_{dk}(t)$ and single Kinect $V_{sk}(t)$ to the spirometer $V_s(t)$– Red labelled keypoints have been incorrectly computed in $V_{sk}(t)$. $V_{sk}(t)$ in (b) has been incorrectly calibrated due to incorrect computation of keypoints caused by the trunk motion.

temporally, and (c) segment the volume–time curve into *tidal volume* and *main effort*, we chose the cut-off frequency as 1 Hz, given the wide range of respiratory rates for adults and elderly at 12 – 36 breaths/minute (0.2 – 0.6 Hz) [28]. However, to avoid volume–time curve over-smoothing, especially at the *main effort* part where the curve slope is critical and needs to be preserved, we increase the cut-off frequency to 3 Hz and filter the original volume-time curve for computing just the PFT measures. Fig. 3 presents the retrieved volume–time curve $V_{dk}(t)$ and its corresponding $V_{sk}(t)$, and their comparison to the volume–time curve $V_s(t)$ obtained from the spirometer, for FVC and SVC tests, respectively. As seen, the trunk motion artifacts have been significantly reduced in the retrieved volume–time curve $V_{dk}(t)$, obtained by the proposed method.

C. Volume-Time Data Analysis

Since depth-based volume–time data presents the subject's trunk volume variations, which is a proxy for the exchanged amount of air within the lungs instead of the real amount of exchanged air, it must be calibrated in order to compute PFT measures correctly. This calibration is performed by linearly scaling the y-axis of the depth-based volume–time data using a scaling factor. Since scaling factors are subject-specific (intra-subject), they are learnt during a training phase for each subject by performing a linear regression analysis between Kinect and spirometer volume–time training data. The main step towards this is to compute keypoints in the volume–time data.

Keypoints computation – Multiple keypoints are automatically identified from the Kinect and spirometer volume–time data by performing an elaborate extrema analysis, as detailed in [6], using the same values for parameters and thresholds. We categorise these keypoints based on their application through the rest of the paper, as follows:

1) Identifying *tidal volume* using $\{\mathcal{C}, \mathcal{D}\}$ and *main effort* using $\{\mathcal{E}, \mathcal{A}, \mathcal{B}\}$.
2) Computing *main effort* scaling factors using $\{\mathcal{A}, \mathcal{B}\}$.

TABLE I
PFT MEASURES OF FVC AND SVC TESTS, THEIR DESCRIPTION AND COMPUTATION METHOD

FVC Measures		SVC Measures	
Computation	**Description**	**Computation**	**Description**
$FVC = V(t_{\mathcal{A}}) - V(t_{\mathcal{B}})$	Maximum amount of air blown out after a maximal inhalation	$VC = V(t_{\mathcal{B}}) - V(t_{\mathcal{A}})$	Volume change between full inhalation and full exhalation
$FEV1 = [V(t_0+1) - V(t_0)]$	Volume of air forcibly exhaled in 1 second starting from 'time zero'	$IC = V(t_{\mathcal{B}}) - \frac{1}{4}\sum_{i=1}^{4} V(t_{\mathcal{G}i})$	Volume change between a slow, full inhalation and passive end-tidal exhalation
$PEF = \dot{V}(t_{PF})$	Maximum speed of exhaled air	$TV = \frac{1}{4}\sum_{i=1}^{4}\left[V(t_{\mathcal{F}i}) - V(t_{\mathcal{G}i})\right]$	Volume of air inhaled and exhaled at rest condition
$FEF_{25\%} = \dot{V}(t_{0.25FVC})$	Flow of exhaled air at 25% of *FVC*	$ERV = \frac{1}{4}\sum_{i=1}^{4}\left[V(t_{\mathcal{G}i}) - V(t_{\mathcal{A}})\right]$	Volume change between passive end-tidal exhalation and complete exhalation
$FEF_{50\%} = \dot{V}(t_{0.50FVC})$	Flow of exhaled air at 50% of *FVC*		
$FEF_{75\%} = \dot{V}(t_{0.75FVC})$	Flow of exhaled air at 75% of *FVC*		
$FEF_{25\text{-}75\%} = \frac{0.75FVC - 0.25FVC}{t(FEF_{25\%}) - t(FEF_{75\%})}$	Mean forced expiratory flow between 25% and 75% of the *FVC*		

3) Computing *tidal volume* and *main effort* similarity measures using $\{\mathcal{F}_i, \mathcal{G}_i\}_{i=1}^{4}$ and $\{\mathcal{A}, \mathcal{B}\}$.
4) Computing PFT measures using $\{\mathcal{E}, \mathcal{A}, \mathcal{B}\}$, $\{\mathcal{F}_i, \mathcal{G}_i\}_{i=1}^{4}$, 'time zero' t_0 and 'Peak Flow' t_{PF}.

Fig. 4 illustrates the computed keypoints for FVC and SVC curves $V_s(t)$, $V_{dk}(t)$ and $V_{sk}(t)$. As shown in Fig. 4(a) and (b), all keypoints are computed correctly for the dual Kinect curve $V_{dk}(t)$ and match their corresponding ones in the spirometer curve $V_s(t)$. However, for the single Kinect curve $V_{sk}(t)$, several keypoints, i.e., $\{\mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}\}$ and $\{\mathcal{F}_i, \mathcal{G}_i\}_{i=1}^{4}$, (labelled red in Fig. 4(a) and (b)), are computed incorrectly due to the effects of the subject's body movement on $V_{sk}(t)$. For example, $V_{sk}(t)$ is not calibrated correctly in Fig. 4(b) because keypoint $\mathcal{B}$ is computed incorrectly, whilst $V_{dk}(t)$ is calibrated quite precisely for the same sequence.

Linear regression analysis – Linear regression is performed separately for *tidal volume* and *main effort* Kinect and spirometer volume–time curves, and provides individual *tidal volume* and *main effort* scaling factors. In order to perform the linear regression, corresponding data samples of the Kinect and spirometer volume–time curves must be identified. Thus, spirometer volume–time data is sampled at the Kinect sampling rate of 30 Hz, and the Kinect and spirometer *tidal volume* are separated using $\{\mathcal{C}, \mathcal{D}\}$ keypoints. *Tidal volume* data are then detrended (see the trend in $V_{sk}(t)$ in Fig. 4(a) and (b)) by applying empirical mode decomposition (EMD) [29]. This increases the similarity between the Kinect and spirometer *tidal volume* curves and attains better temporal alignment. Finally, the delay is computed using a windowed cross correlation between these curves and used to temporally align the whole Kinect and spirometer volume–time data. This process is carried out for $V_{dk}(t)$ and $V_{sk}(t)$ separately.

The *tidal volume* and *main effort* scaling factors are computed by establishing linear regression for *tidal volume* as

$$\widehat{V}_s^{tv} = \xi_{dk}^{tv} \cdot \widehat{V}_{dk}^{tv} + \psi_{dk}^{tv}, \quad \widehat{V}_{dk}^{tv} = \widehat{V}_{dk}(t)\Big|_{t_C}^{t_{\mathcal{D}}}, \tag{1}$$

and for *main effort* individually as

$$\left\langle \widehat{V}_s(t_{\mathcal{A}}), \widehat{V}_s(t_{\mathcal{B}}) \right\rangle = \xi_{dk}^{me} \cdot \left\langle \widehat{V}_{dk}(t_{\mathcal{A}}), \widehat{V}_{dk}(t_{\mathcal{B}}) \right\rangle + \psi_{dk}^{me}, \tag{2}$$

where $\widehat{V}_{dk}(t)$ and $\widehat{V}_s(t)$ are detrended and zero mean normalised volume–time data of Kinect and spirometer. Since volume–time data are normalised to zero mean of their *tidal volume*, then $\psi_{dk}^{tv} \approx 0$. Thus, the *tidal volume* and *main effort* scaling factors are defined as $\langle \xi_{dk}^{tv} \rangle$ and $\langle \xi_{dk}^{me}, \psi_{dk}^{me} \rangle$, respectively. Similarly, the *tidal volume* and *main effort* scaling factors, i.e., $\langle \xi_{sk}^{tv} \rangle$ and $\langle \xi_{sk}^{me}, \psi_{sk}^{me} \rangle$, are computed from the single Kinect volume–time data $V_{sk}(t)$ for comparative analysis.

D. PFT Measures Computation

Within a spirometry test, several clinical PFT measures are provided by the spirometer software. Besides these numerical measures, pulmonologists often use volume–time (for FVC and SVC tests) and flow–volume (for FVC only) spirograms [6] as a qualitative presentation of lung function. Here, we compute seven primary FVC measures and all of four SVC measures, from depth-based volume–time and flow–time data, using the required keypoints. Table I presents all PFT measures, their description and computation method. The groundtruth measures were obtained directly from the spirometer software, for evaluation and comparison.

E. Learning Intra-Subject Scaling Factors

The aim of this study is to assess human lung function remotely and independently, without support from any clinical device, e.g., a spirometer. The coefficients of the linear regression, i.e., the scaling factors, between trunk volume and lungs air flow, are subject-specific and depend on physical body specifications, e.g., weight, height, BMI, gender and race. Thus, we train our system to learn the scaling factors per subject (intra-subject), which enables it to perform a PFT test independent of a spirometer at later trials.[2]

In the training phase, intra-subject scaling factors are learnt using Kinect and spirometer training trials, and computed as $\left\{\langle \xi_{dk}^{tv} \rangle^{\ell}\right\}_{\ell=1}^{n_{tv}}$ & $\left\{\langle \xi_{dk}^{me}, \psi_{dk}^{me} \rangle^{\ell}\right\}_{\ell=1}^{n_{me}}$ for $\left\{\left\langle V_{dk}(t), V_s(t)\right\rangle^{\ell}\right\}_{\ell=1}^{n_T}$ as explained in Section III-C, where n_{tv} and n_{me} are number of *tidal volume* and *main effort* training trials, and $n_T = n_{tv} + n_{me}$.

In the testing phase, first, the depth-based volume–time data of a test trial, i.e., $V_{dk}^{test}(t)$, is retrieved using the proposed method, explained in Section III-A. Then, *tidal volume* and

[2]A trial refers to each performance of the FVC/SVC test by each subject.

TABLE II, III

STATISTICAL COMPARISON OF MEAN AND STANDARD DEVIATION OF L_2 ERROR (μ_{dk} & μ_{sk} AND σ_{dk} & σ_{sk}), RATIO OF MEAN OF L_2 ERROR TO THE MEAN VALUE OF THAT MEASURE (Ω_{dk} & Ω_{sk}), AND CORRELATION COEFFICIENTS (λ_{dk} & λ_{sk}) BETWEEN THE DEPTH-BASED (THE DUAL AND SINGLE KINECT APPROACHES) AND THE SPIROMETER MEASURES. ALTHOUGH THE EVALUATION RESULTS OF THE DUAL KINECT METHOD SHOW IMPROVEMENT ACROSS ALL MEASURES, THE BOLD NUMBERS POINT TO THE MEASURES WHERE THEIR ERROR (μ_{dk} & σ_{dk} & Ω_{dk}) HAS REMARKABLY DROPPED BY HALF AND THEIR CORRELATION COEFFICIENTS (λ_{dk}) HAVE SIMILARLY IMPROVED.

TABLE II
FVC MEASURES EVALUATION RESULTS

		FVC	*FEV1*	*PEF*	$FEF_{25\%}$	$FEF_{50\%}$	$FEF_{75\%}$	$FEF_{25\text{-}75\%}$
dual Kinect	μ_{dk}	0.36	0.55	1.83	2.02	**1.55**	**0.85**	**1.28**
	σ_{dk}	0.39	0.55	1.59	1.92	**1.28**	**0.86**	**1.04**
	Ω_{dk}	0.07	0.13	0.22	0.26	**0.32**	**0.39**	**0.31**
	λ_{dk}	0.87	0.68	0.46	0.35	**0.34**	**0.47**	**0.43**
single Kinect	μ_{sk}	0.45	0.68	2.32	2.48	3.29	2.60	2.67
	σ_{sk}	0.40	0.49	2.30	2.45	3.17	2.97	2.48
	Ω_{sk}	0.09	0.17	0.28	0.32	0.68	1.22	0.64
	λ_{sk}	0.86	0.67	0.41	0.22	0.09	0.15	0.19

TABLE III
SVC MEASURES EVALUATION RESULTS

		VC	*IC*	*TV*	*ERV*
dual Kinect	μ_{dk}	0.421	**0.393**	0.148	**0.315**
	σ_{dk}	0.474	**0.409**	0.150	**0.331**
	Ω_{dk}	0.095	**0.127**	0.160	**0.239**
	λ_{dk}	0.869	**0.810**	0.839	**0.748**
single Kinect	μ_{sk}	0.550	0.730	0.181	0.778
	σ_{sk}	0.656	0.712	0.202	0.812
	Ω_{sk}	0.124	0.235	0.196	0.592
	λ_{sk}	0.719	0.466	0.767	0.389

main effort similarity measures

$$F_{tv} = \frac{1}{4}\sum_{i=1}^{4}\Big[V_{dk}(t_{\mathcal{F}i}) - V_{dk}(t_{\mathcal{G}i})\Big], \tag{3}$$

$$F_{me} = \Big[V_{dk}(t_{\mathcal{B}}) - V_{dk}(t_{\mathcal{A}})\Big], \tag{4}$$

are computed as F_{tv}^{test} & F_{me}^{test} and $\{F_{tv}^{\ell}\}_{\ell=1}^{n_{tv}}$ & $\{F_{me}^{\ell}\}_{\ell=1}^{n_{me}}$ for $V_{dk}^{test}(t)$ and $\{\langle V_{dk}(t)\rangle^{\ell}\}_{\ell=1}^{n_T}$, respectively. These allow for optimisation of *tidal volume* and *main effort* training trials by matching training similarity measures with the similarity measures of $V_{dk}^{test}(t)$:

$$J_{tv} = \underset{j\in[1..n_{tv}]}{\arg\min}\Big\{\big|F_{tv}^{test} - F_{tv}^{j}\big|\Big\}, \tag{5}$$

$$J_{me} = \underset{j\in[1..n_{me}]}{\arg\min}\Big\{\big|F_{me}^{test} - F_{me}^{j}\big|\Big\}. \tag{6}$$

The associated scaling factors of J_{tv} and J_{me} trials, declared as $\langle\xi_{dk}^{tv}\rangle^{J_{tv}}$ and $\langle\xi_{dk}^{me}, \psi_{dk}^{me}\rangle^{J_{me}}$ are then used to calibrate $V_{dk}^{test}(t)$ as

$$V_{dk}^{cal}(t) = \Big[V_{dk}^{test}(t)\cdot\langle\xi_{dk}^{tv}\rangle^{J_{tv}}\Big]_{t=t_{\mathcal{C}}}^{t=t_{\mathcal{D}}} + \Big[V_{dk}^{test}(t)\cdot\langle\xi_{dk}^{me}\rangle^{J_{me}} + \langle\psi_{dk}^{me}\rangle^{J_{me}}\Big]_{t=t_{\mathcal{D}}}^{t=max(t_{\mathcal{A}},t_{\mathcal{B}})}. \tag{7}$$

In order to compare our method to the single Kinect approach, a similar process is carried out to obtain $\langle\xi_{sk}^{tv}\rangle^{J'_{tv}}$ and $\langle\xi_{sk}^{me}, \psi_{sk}^{me}\rangle^{J'_{me}}$ and calibrate $V_{sk}^{test}(t)$, where J'_{tv} and J'_{me} are the optimised *tidal volume* and *main effort* selected trials.

We evaluated the intra-subject training and testing process using leave-one-out cross-validation, which is the most suitable validation method for our approach, due to the limited number of trials for each subject. Thus, for each subject, one trial is repeatedly considered as the *test* and the model is trained with the rest of the trials.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. System Configuration and Dataset Specification

We acquired the depth data using two facing Kinect V2 sensors, with the subject sitting in between, as shown in Fig. 1. Each of the Kinects was placed at the distance of ∼1.5 m away from the subject to minimise the noise [6], [27] and at a height of 0.6 m. Each subject was asked to wear a reasonably tight T-shirt and sit up straight on a backless chair. Participants were neither restricted nor advised to be stationary during the PFTs, and the tests were performed as routine spirometry.

Thirty five subjects (8 females and 27 males) of various ages (30.3 ± 5.3) and BMIs (23.9 ± 3.1) participated in this study. Ethical approval was obtained from the University of Bristol Research Ethics Committee (Reference 56124), and each participant signed a written consent form. According to the spirometry experiment protocols [2], each subject must perform several FVC and SVC tests (at least three times) to achieve consistent PFT measures. Thus, most of the subjects had to perform extra tests to ensure consistency.

A total of 306 PFT sessions were held, of which only 8 sessions' data (8 sequences) were dropped. The data for five sessions were omitted due to the spirometer (two sessions) and Kinect (three sessions) failure, and one session's data was removed as a subject occluded the chest by hands during the test. There were only two sequences which the proposed method failed to compute their keypoints due to complex body motion patterns. Otherwise, volume–time data of all the other 298 sequences were successfully retrieved and analysed, and their PFT measures were computed and considered in the experimental analysis.

B. PFT Measures Evaluation

Tables II and III present the results of PFT measures for all 35 subjects, computed for 155 FVC and 143 SVC sequences respectively, from $V_{dk}(t)$ & $V_{sk}(t)$. These Tables report, (i) mean and standard deviation of L_2 error (μ_{dk} & μ_{sk} and σ_{dk} & σ_{sk}) for each measure, (ii) ratio of mean of L_2 error to the mean value of that measure (Ω_{dk} & Ω_{sk}), and (iii) correlation coefficients (λ_{dk} & λ_{sk}) between the depth-based and the spirometer measures.

As can be seen in Table II, (μ_{dk}, σ_{dk}, Ω_{dk}) have decreased for the dual Kinect approach across all measures, compared to their single Kinect [5], [6] counterparts (μ_{sk}, σ_{sk}, Ω_{sk}). In particular, these errors have dropped by half for $FEF_{50\%}$, $FEF_{75\%}$ and

TABLE IV
RESULTS OF BLAND-ALTMAN ANALYSIS OF THE DUAL AND SINGLE FVC AND SVC MEASURES

	FVC Measures							SVC Measures			
	FVC	*FEV1*	*PEF*	$FEF_{25\%}$	$FEF_{50\%}$	$FEF_{75\%}$	$FEF_{25-75\%}$	*VC*	*IC*	*TV*	*ERV*
$\mathcal{L}_{dk} - \mathcal{U}_{dk}$	-0.9–1.1	-1.0–1.7	-3.5–5.3	-3.9–6.1	**-3.4–4.2**	**-2.1–2.5**	**-2.2–3.6**	-1.4–0.8	**-1.2–0.7**	-0.4–0.4	**-0.9–0.8**
$\mathcal{L}_{sk} - \mathcal{U}_{sk}$	-1.1–1.2	-1.6–1.7	-7.0–5.5	-6.0–7.4	**-10.0–6.0**	**-8.6–5.5**	**-7.9–5.7**	-1.8–1.5	**-2.1–1.9**	-0.4–0.6	**-2.2–2.1**
$\mathcal{M}_{dk}$ (%)	98.0%	95.5%	94.2%	93.5%	**95.5%**	**94.2%**	**92.9%**	95.1%	**95.8%**	93.0%	**95.1%**
$\mathcal{M}_{sk}$ (%)	92.2%	92.2%	83.2%	85.1%	**70.3%**	**67.7%**	**63.8%**	90.2%	**76.2%**	88.1%	**64.3%**

Note: $\mathcal{L}_{dk}$ & $\mathcal{U}_{dk}$ and $\mathcal{L}_{sk}$ & $\mathcal{U}_{sk}$ indicate the *lower & upper limits of agreement* for the dual and the single Kinect PFT measures, respectively. $\mathcal{M}_{dk}$ and $\mathcal{M}_{sk}$ state the percentage of trials where the difference between the dual Kinect and the single Kinect measures with the spirometer measure lies in $\mathcal{L}_{dk} - \mathcal{U}_{dk}$.

$FEF_{25-75\%}$ measures. This remarkable error reduction is due to the measures being computed using the top curvature of the *main effort* of volume–time data, which was successfully recovered in $V_{dk}(t)$ by decoupling trunk movements from the respiratory motion (compare in Fig. 3 against $V_{sk}(t)$).

For the other measures reported in Table II, $(\mu_{dk}, \sigma_{dk}, \Omega_{dk})$ have not decreased significantly compared to their single Kinect [5], [6] counterparts. For *FVC*, this is because the measure is computed using the same keypoints $\mathcal{A}$ and $\mathcal{B}$, which are also used in the *main effort* calibration. *FEV1*, *PEF* and $FEF_{25\%}$ measures are computed from the steepest part of the *main effort*, between keypoints t_0 and $t_{FEF25\%}$. Thus, we believe the trunk forward movement at the start of forceful exhalation increases the *main effort* curve slope and accidentally contributes in achieving better *FEV1*, *PEF* and $FEF_{25\%}$ measures.

These results confirm the superiority of the proposed method to the single Kinect approach [5], [6], with λ_{dk} also showing better correlation of PFT measures than λ_{sk}. However, λ_{dk} does not express strong correlation between the dual-Kinect-based FVC measures and the spirometer, except for the *FVC* and *FEV1*. This is expected as we exploited all of acquired data and did not remove the trials that impose high error. In particular, these trials appear as outliers and influence the correlation coefficients. To further clarify this issue, we have performed a Bland-Altman analysis of PFT measures (Section IV-C) and present more qualitative and quantitative comparison between depth-based and spirometer PFT measures.

Ostadabbas *et al.* [7] reported a 0.88 average correlation with a spirometer for *FEV1* (and no other measure). However, this cannot be directly compared to the *FEV1* correlation coefficient computed here which is on a different dataset, acquired by different protocols, under different criteria.

Table III reports the evaluation results for SVC measures, in which $(\mu_{dk}, \sigma_{dk}, \Omega_{dk})$ have also dropped by half for *IC* and *ERV* measures, compared to $(\mu_{sk}, \sigma_{sk}, \Omega_{sk})$. Moreover, λ_{dk} shows much better correlation for these two measures, compared to λ_{sk}. The improved results are due to the trunk motion corrections, which have removed the offset between the *tidal volume* and the *main effort*. The *VC* measure was computed using the keypoints $\mathcal{A}$ and $\mathcal{B}$, which were also exploited for calibrating SVC volume–time data, thus the proposed method achieved only a slight improvement in this measure. *TV* was also slightly improved as subjects' movements in the rest condition is insignificant.

PFT measures' correlation coefficient and error, reported in [5], [6], are relatively better than the results reported here because [5], [6] were evaluated on a dataset in which the subject's trunk motion were strictly restrained during the test. Tables II and III report the results of applying the same single Kinect method in [5], [6] on the current dataset in which subjects performed PFT as a routine spirometry test and their body's normal reaction to deep and forced *inhalation-exhalation* was not restricted. Comparing the evaluation results obtained from the dual Kinect approach to the single Kinect method on this dataset (see Tables II and III), confirms that eliminating trunk motion, achieved by the dual Kinect approach, highly improves the PFT measures' correlation and reduces the error, even when both approaches use the same volume–time data analysis and intra-subject scaling factor learning methods.

C. Bland-Altman Analysis of PFT Measures

Table IV reports Bland-Altamn [30] range of agreement between the dual Kinect and the spirometer measures, i.e., $\mathcal{L}_{dk} - \mathcal{U}_{dk}$, and also between the single Kinect and the spirometer measures, i.e., $\mathcal{L}_{sk} - \mathcal{U}_{sk}$, where $\mathcal{L}_{dk}$ & $\mathcal{U}_{dk}$ and $\mathcal{L}_{sk}$ & $\mathcal{U}_{sk}$ indicate the *lower & upper limits of agreement* for the dual and the single Kinect measures, respectively. Results confirms that the dual Kinect measures better agree with the spirometer across all the measures, particularly for $FEF_{50\%}$, $FEF_{75\%}$, $FEF_{25-75\%}$, *IC* and *ERV*.

Further, in order to better compare the error between the dual and single Kinect PFT measures, $\mathcal{M}_{dk}$ was computed as the percentage of trials where the difference between the dual Kinect measure and the spirometer measure lies in $\mathcal{L}_{dk} - \mathcal{U}_{dk}$. Similarly, $\mathcal{M}_{sk}$ specifies the percentage of trials in the same range of agreement between the single Kinect measure and the spirometer measure (see Table IV). Although $\mathcal{M}_{dk}$ is greater than $\mathcal{M}_{sk}$ across all PFT measures, the difference between $\mathcal{M}_{dk}$ and $\mathcal{M}_{sk}$ is more distinguishable for $FEF_{50\%}$, $FEF_{75\%}$, $FEF_{25-75\%}$, *IC* and *ERV*. Fig. 5 shows Bland-Altman plots of $FEF_{75\%}$, $FEF_{25-75\%}$ and *ERV* measures for the dual and single Kinect approaches.

D. Performance Evaluation of Similarity Measures

We evaluated the performance of the *tidal volume* and *main effort* similarity measures (3) and (4), in terms of their ability to choose the intra-subject scaling factors $\langle \xi_{dk}^{tv} \rangle^{J_{tv}}$ & $\langle \xi_{dk}^{me} \rangle^{J_{me}}$, which are supposed to calibrate the test volume–time data with the minimum error, among the training scaling factors $\left\{ \langle \xi_{dk}^{tv} \rangle^{\ell} \right\}_{\ell=1}^{n_{tv}}$ & $\left\{ \langle \xi_{dk}^{me} \rangle^{\ell} \right\}_{\ell=1}^{n_{me}}$. Thus, we used normalised L_2 error $\mathtt{SME}_{dk}^{tv}$ & $\mathtt{SME}_{dk}^{me}$, computed as the ratio of L_2 error between $\langle \xi_{dk}^{tv} \rangle^{J_{tv}}$ & $\langle \xi_{dk}^{me} \rangle^{J_{me}}$ and $\langle \xi_{dk}^{tv} \rangle^{c}$ & $\langle \xi_{dk}^{me} \rangle^{c}$, to $\langle \xi_{dk}^{tv} \rangle^{c}$ & $\langle \xi_{dk}^{me} \rangle^{c}$.

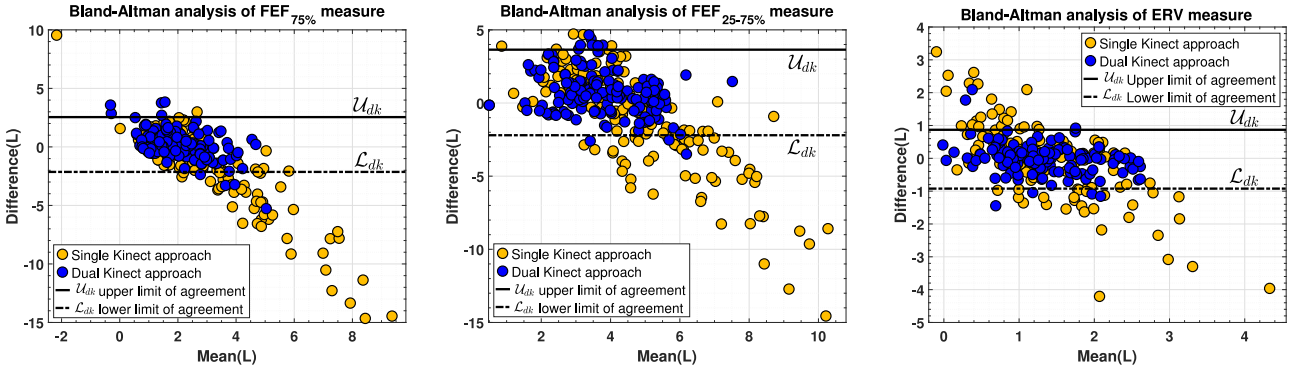


Fig. 5. Bland-Altman plots for $FEF_{75\%}$ (a), $FEF_{25-75\%}$ (b) and *ERV* (c), measures. Many of the single Kinect PFT measures lie outside of the lower limits of agreement, i.e., $\mathcal{L}_{dk}$, and the upper limit of agreement, i.e., $\mathcal{U}_{dk}$, computed for the dual Kinect PFT measures.

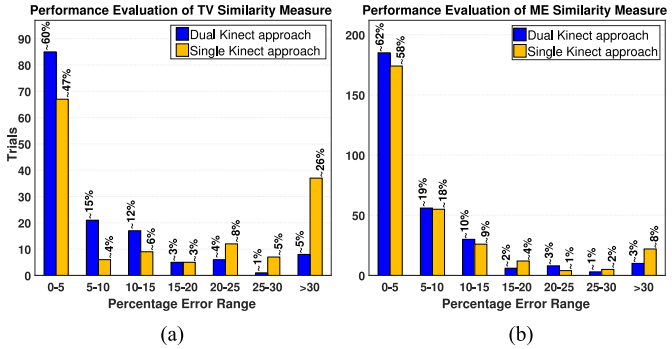


Fig. 6. Performance evaluation of similarity measures by distributing 143 *tidal volume* trials (a), and 298 *main effort* trials (b) over $\mathtt{SME}^{tv}_{dk}$ & $\mathtt{SME}^{tv}_{sk}$ and $\mathtt{SME}^{me}_{dk}$ & $\mathtt{SME}^{me}_{sk}$ at various intervals for the dual (blue) and single (orange) Kinect approaches.

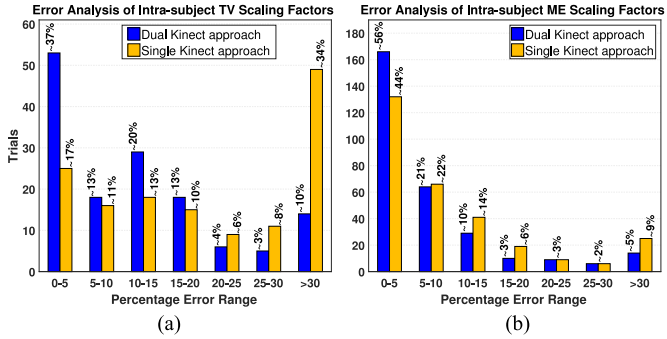


Fig. 7. Error analysis of intra-subject scaling factors by distributing 143 *tidal volume* trials (a), and 298 *main effort* trials (b) over $\mathtt{SCE}^{tv}_{dk}$ & $\mathtt{SCE}^{tv}_{sk}$ and $\mathtt{SCE}^{me}_{dk}$ & $\mathtt{SCE}^{me}_{sk}$ at various intervals for the dual (blue) and single (orange) Kinect approaches.

$\langle \xi^{tv}_{dk} \rangle^c$ & $\langle \xi^{me}_{dk} \rangle^c$ are the numerically closest scaling factors to the spirometer scaling factors of the test trial, i.e., $\langle \xi^{tv}_{dk} \rangle^o$ & $\langle \xi^{me}_{dk} \rangle^o$. Note that, $\langle \xi^{tv}_{dk} \rangle^o$ & $\langle \xi^{me}_{dk} \rangle^o$ are computed using the spirometer volume–time data of the test trial and are only used for evaluation and comparison. Similarly, $\mathtt{SME}^{tv}_{sk}$ and $\mathtt{SME}^{me}_{sk}$ are also computed for the single Kinect approach [5], [6]. Fig. 6(a) and (b) show the distribution of *tidal volume* and *main effort* trials over the computed error $\mathtt{SME}^{tv}_{dk}$ & $\mathtt{SME}^{tv}_{sk}$ and $\mathtt{SME}^{me}_{dk}$ & $\mathtt{SME}^{me}_{sk}$ for the dual (blue) and the single (orange) Kinect approaches, respectively, in the range $0-30\%$ at 5% intervals and then for more than 30%. As can be seen, $\sim$75% of *tidal volume* trials and $\sim$81% of *main effort* trials in the dual Kinect approach have $<10\%$ error. This reduces to $\sim$51% and $\sim$76% in the single Kinect approach. Also, many fewer trials with $>30\%$ error occur in the dual Kinect approach, i.e., $\sim$5%, as opposed to $\sim$26% in the single Kinect approach.

E. Error Analysis of Intra-Subject Scaling Factors

We obtain the spirometer scaling factors $\langle \xi^{tv}_{dk} \rangle^o$ & $\langle \xi^{me}_{dk} \rangle^o$ to assist us in evaluating our intra-subject scaling factors by computing the normalised L_2 error, i.e., $\mathtt{SCE}^{tv}_{dk}$ and $\mathtt{SCE}^{me}_{dk}$, between $\langle \xi^{tv}_{dk} \rangle^{J_{tv}}$ & $\langle \xi^{me}_{dk} \rangle^{J_{me}}$ and $\langle \xi^{tv}_{dk} \rangle^o$ & $\langle \xi^{me}_{dk} \rangle^o$. We also compare against the single Kinect approach [5], [6] by computing $\mathtt{SCE}^{tv}_{sk}$ and $\mathtt{SCE}^{me}_{sk}$. Fig. 7(a) and (b) present the distribution of *tidal volume* and *main effort* trials over the intra-subject scaling factor errors $\mathtt{SCE}^{tv}_{dk}$ & $\mathtt{SCE}^{tv}_{sk}$ and $\mathtt{SCE}^{me}_{dk}$ & $\mathtt{SCE}^{me}_{sk}$ for the dual Kinect (blue) and the single Kinect (orange) approaches, respectively. For example, in Fig. 7(a), $\sim$50% of *tidal volume* trials have $<10\%$ error in the dual Kinect approach against $\sim$28% in the single Kinect approach. Also, only $\sim$10% of the *tidal volume* trials have $>30\%$ error for the dual Kinect against $\sim$34% in the single Kinect. In the *main effort* trials, the dual Kinect approach similarly performs better (see Fig. 7(b)).

F. Statistical Analysis of Within-Subject Scaling Factors

Table V reports the mean and standard deviation of within-subject *tidal volume* and *main effort* scaling factors, for the dual and single Kinect [5], [6] approaches for all 35 participants, denoted as $\mathtt{M}^{tv}_{dk}$, $\mathtt{M}^{me}_{dk}$ & Σ^{tv}_{dk}, Σ^{me}_{dk} and $\mathtt{M}^{tv}_{sk}$, $\mathtt{M}^{me}_{sk}$ & Σ^{tv}_{sk}, Σ^{me}_{sk}, respectively. Minimum to maximum range of scaling factors and their distribution between the 1st and 3rd quartiles along with the outliers are presented in Fig. 8.

The comparison between the scaling factors standard deviation, i.e., Σ^{tv}_{dk} & Σ^{me}_{dk} versus Σ^{tv}_{sk} & Σ^{me}_{sk} in Table V, shows that dual Kinect within-subject scaling factors are more consistent than the single Kinect method [5], [6], especially for the *tidal volume* scaling factors. This can be better realised by com-

TABLE V

MEAN AND STANDARD DEVIATION OF WITHIN-SUBJECT *tidal volume* AND *main effort* SCALING FACTORS FOR THE DUAL KINECT APPROACH, I.E., $\mathtt{M}^{tv}_{dk}$, $\mathtt{M}^{me}_{dk}$ & Σ^{tv}_{dk}, Σ^{me}_{dk}, AND THE SINGLE KINECT APPROACH, I.E, $\mathtt{M}^{tv}_{sk}$, $\mathtt{M}^{me}_{sk}$ & Σ^{tv}_{sk}, Σ^{me}_{sk}. BLUE NUMBERS HIGHLIGHT SOME SAMPLES OF THE BEST AGREEMENT BETWEEN $\mathtt{M}^{tv}_{dk}$ AND $\mathtt{M}^{me}_{dk}$ WHEREAS THEIR CORRESPONDING VALUES IN THE SINGLE KINECT APPROACH, I.E., $\mathtt{M}^{tv}_{sk}$ AND $\mathtt{M}^{me}_{sk}$, DISAGREE WITH EACH OTHER. **BOLD RED** NUMBERS ARE RELATED TO Σ^{tv}_{sk} OR Σ^{me}_{sk}, WHERE THEIR VALUES ARE CONSIDERABLY HIGHER THAN THE CORRESPONDING Σ^{tv}_{dk} AND Σ^{me}_{dk}. RED NUMBERS HIGHLIGHT THE Σ^{tv}_{dk} OR Σ^{me}_{dk} WHERE THEIR VALUE ARE HIGHER THAN VALUE OF Σ^{tv}_{sk} AND Σ^{me}_{sk}.

		S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	S29	S30	S31	S32	S33	S34	S35
dual Kinect	$\mathtt{M}^{tv}_{dk}$	4.82	6.81	3.19	3.29	9.05	5.95	6.33	**6.47**	3.08	4.51	7.74	4.06	5.69	5.53	4.78	4.23	5.14	5.25	6.92	5.47	10.2	5.63	7.99	**5.71**	**5.22**	5.90	5.59	4.52	5.95	4.36	6.34	3.27	7.14	6.07	5.28
	$\mathtt{M}^{me}_{dk}$	4.72	5.73	3.40	3.30	7.18	5.65	5.75	**6.56**	2.97	4.30	4.61	3.77	4.93	5.74	4.18	5.32	4.17	3.62	6.10	5.01	5.17	4.42	5.29	**5.45**	**5.16**	4.89	5.33	5.00	4.97	3.91	3.92	3.04	5.42	4.31	7.91
	Σ^{tv}_{dk}	0.57	1.03	0.49	0.48	0.55	0.36	0.30	1.01	0.55	0.32	0.32	0.19	0.28	0.85	0.69	1.51	0.54	0.42	0.64	0.58	2.61	0.76	0.31	1.30	0.35	0.54	0.78	1.84	0.24	0.33	1.31	0.16	0.72	0.54	3.48
	Σ^{me}_{dk}	0.41	0.29	0.56	0.43	0.47	0.48	0.73	0.62	0.23	0.40	0.27	0.94	0.19	1.21	0.17	0.61	0.52	0.30	0.27	0.50	1.24	0.25	0.48	1.29	0.22	0.43	0.62	0.46	0.19	0.45	0.10	0.17	1.23	0.61	1.75
Single Kinect	$\mathtt{M}^{tv}_{sk}$	4.48	9.05	2.71	3.32	9.30	6.88	5.93	**2.98**	3.09	4.30	6.71	4.16	5.54	5.25	4.69	5.95	5.58	5.39	5.53	5.59	7.03	5.88	8.81	**4.96**	**7.14**	7.67	5.72	4.95	8.41	3.95	7.78	3.43	9.28	6.24	3.76
	$\mathtt{M}^{me}_{sk}$	4.67	5.79	3.43	3.48	7.11	5.59	5.84	**12.5**	3.00	4.34	4.86	3.64	4.91	6.08	4.25	6.34	5.32	3.56	6.11	5.59	5.14	4.45	5.29	**8.82**	**11.9**	5.18	5.31	5.21	4.92	3.92	4.04	3.42	5.88	4.30	7.86
	Σ^{tv}_{sk}	1.11	**2.34**	0.90	0.66	1.01	0.39	0.40	1.05	**1.46**	0.65	**2.68**	**0.77**	0.53	**1.82**	0.52	**3.47**	**1.73**	0.45	**2.55**	0.92	2.76	0.99	**1.28**	1.48	**1.32**	1.11	0.73	**2.27**	**1.43**	0.75	**2.70**	0.35	**1.69**	1.05	1.87
	Σ^{me}_{sk}	0.46	0.54	0.61	0.69	0.40	0.51	0.74	**4.31**	0.27	0.57	0.68	0.67	0.16	**2.38**	0.29	**2.03**	**1.94**	0.24	0.21	**1.07**	1.30	0.30	0.46	**6.81**	**6.16**	0.70	0.60	0.84	0.21	0.47	0.36	**0.76**	1.56	0.62	1.62

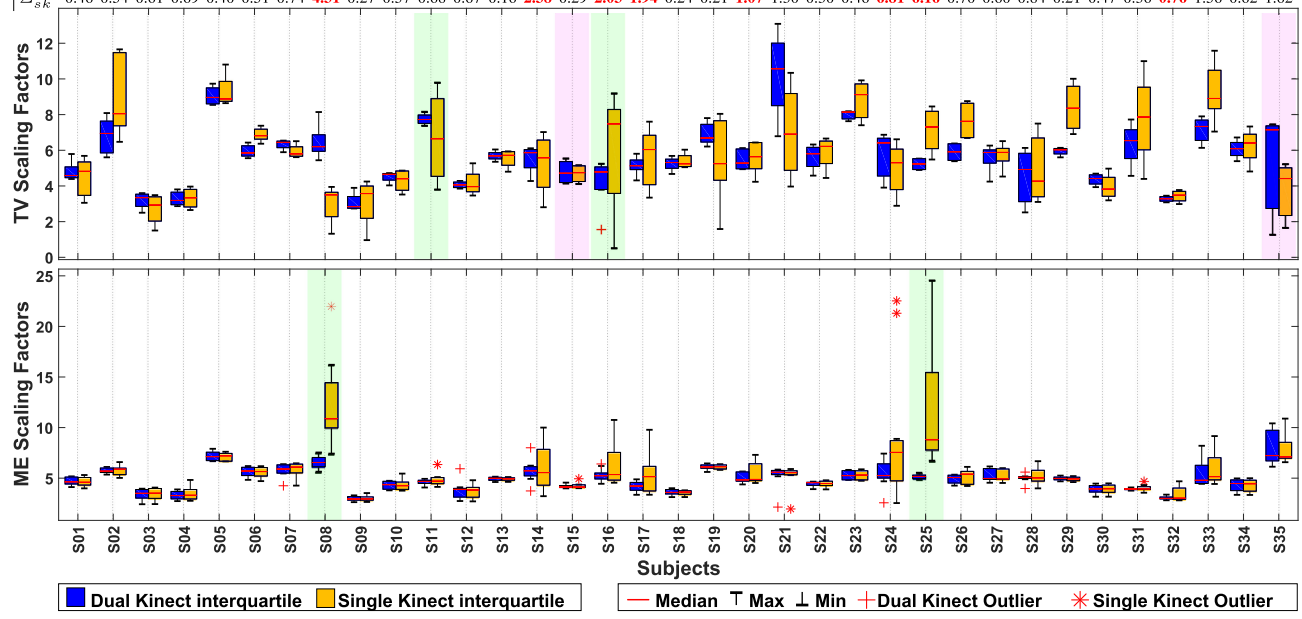


Fig. 8. Boxplot of statistics of within-subject *tidal volume* (top) and *main effort* (bottom) scaling factors of 35 subjects' trials, in which interquartile range, median, max, min and outliers of *tidal volume* and *main effort* scaling factors are illustrated for the dual (blue) and single (orange) Kinect approaches. The interquartile range of the single Kinect *tidal volume* scaling factors are wider across all subjects except for a few, e.g., subjects 15 and 35 (pink highlighted). In particular, Σ^{tv}_{sk} is 8.4 and 2.3 times higher than Σ^{tv}_{dk} for the green highlighted subjects 11 and 16. Similarly, for the *main effort* scaling factors, Σ^{me}_{sk} is 6.9 and 28 times higher than Σ^{me}_{dk} for the green highlighted subjects 8 and 25.

paring min to max range of the scaling factors, and also their interquartile ranges in Fig. 8, for the dual Kinect (blue boxes) and the single Kinect (orange boxes) approaches. Among these, only subject 35 has a considerably greater Σ^{tv}_{dk} (red) than Σ^{tv}_{sk}, whereas Σ^{tv}_{sk}, Σ^{me}_{sk} are higher for numerous subjects (bold red). For example, Σ^{tv}_{sk} is 8.4 and 2.3 times higher than Σ^{tv}_{dk} for subjects 11 and 16, and Σ^{me}_{sk} is 6.9 and 28 times higher than Σ^{me}_{dk} for subjects 8 and 25 (highlighted in green in Fig. 8). The greater the scaling factors' standard deviation is, the higher the depth-based PFT measures' error would be. For example, the average error of *TV* and *FVC* measures decreases from 0.23 and 0.84 in the single Kinect approach to 0.07 and 0.19 in the dual Kinect approach for subjects 16 and 25, respectively.

Finally, Table VI shows the mean ($\mu_{\mathtt{M}'}$) and standard deviation ($\sigma_{\mathtt{M}'}$) of the absolute difference between 'the average of within-subject *tidal volume* scaling factors' and 'the average of within-subject *main effort* scaling factors,' i.e., $\mathtt{M}' = |\mathtt{M}^{tv}_x - \mathtt{M}^{me}_x|$, where $x = dk$ for the dual Kinect approach and $x = sk$ for the single Kinect method. It also shows the normalised mean of $\mathtt{M}'$ as

TABLE VI

STATISTICS OF $\mathtt{M}' = |\mathtt{M}^{tv}_x - \mathtt{M}^{me}_x|$ AND $\Sigma' = |\Sigma^{tv}_x - \Sigma^{me}_x|$, x=*sk or dk* ACROSS 35 SUBJECTS IN DUAL AND SINGLE KINECT METHODS

	$\mu_{\mathtt{M}'}$	$\sigma_{\mathtt{M}'}$	$\Omega_{\mathtt{M}'}$	$\mu_{\Sigma'}$	$\sigma_{\Sigma'}$	$\Omega_{\Sigma'}$
Dual Kinect	**1.02**	**1.10**	**0.19**	**0.37**	**0.44**	**0.57**
Single Kinect	1.72	1.98	0.81	1.03	1.28	0.85

$\Omega_{\mathtt{M}'} = \mu_{\mathtt{M}'}/\alpha_{\mathtt{M}'}$, in which the normalisation factor $\alpha_{\mathtt{M}'}$ is defined as the average of $\{\mathtt{M}^{tv}_x, \mathtt{M}^{me}_x\}$ across all subjects. Table VI also presents similar statistics for $\Sigma' = |\Sigma^{tv}_x - \Sigma^{me}_x|$ ($\mu_{\Sigma'}, \sigma_{\Sigma'}, \Omega_{\Sigma'}$). As seen, mean, standard deviation and the normalised mean of $\mathtt{M}'$ & Σ', are notably smaller in the dual Kinect method, where it shows better agreement between *tidal volume* and *main effort* scaling factors. For example, $\mathtt{M}^{tv}_{dk}$ and $\mathtt{M}^{me}_{dk}$ are almost equal for subjects 8, 24 and 25 (in blue in Table V), whereas $\mathtt{M}^{tv}_{sk}$ and $\mathtt{M}^{me}_{sk}$ show a considerable disagreement for these subjects (in orange in Table V).

V. Conclusion and Future Work

We introduced depth-based whole body photoplethysmography to increase remote PFT accuracy by decoupling subject's trunk movements from the respiratory motions using two opposing Kinects. First, two Kinects were calibrated and synchronised to construct a dynamic 3-D model of the subject performing PFT. Using a 3-D mask, thoracoabdominal volume is automatically segmented and used to retrieve a depth-based volume–time data. This volume–time data was then calibrated using the intra-subject scaling factors, learnt in a training phase, and 11 clinical PFT measures were computed. We validated the dPPG PFT measures by comparing them to the measures obtained from a spirometer. The evaluation results show very good improvement compared to the single Kinect approach [5], [6].

The proposed dPPG method does not perform in real-time as the body data acquisition, trunk reconstruction and PFT computation stages operate separately. While the data acquisition and the PFT computation stages perform in nearly real-time, the trunk reconstruction for each PFT performance is accomplished in less than a minute. However, we feel confident to project that by applying GPU-based 3-D reconstruction techniques, and incorporating these stages using further development, dPPG would operate in real-time.

The proposed method for decoupling body movements from respiratory motions results in *tidal volume* and *main effort* scaling factors that are more consistent and better agree with each other (than our earlier method in [5], [6]). However, they are not identical enough to be a unique intra-subject scaling factor that could be used to calibrate the whole volume–time data. We note that in different subjects, thoracoabdominal wall regions contribute differently in the *tidal volume* breathing, and the *main effort inhalation–exhalation*. In our future work, we shall investigate a multi-patch linear regression model to solve this issue.

Acknowledgment

The authors would like to thank the subjects who participated in this research. The authors would also like to thank SPHERE[3], an EPSRC Interdisciplinary Research Centre, for providing the opportunity of collaboration between engineering and clinical researchers. The dPPG PFT dataset is publicly available for download at http://doi.org/ckrh.

References

[1] M. Naghavi *et al.*, "Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013," *Lancet*, vol. 385, no. 9963, pp. 117–171, 2015.

[2] M. Miller *et al.*, "Standardisation of spirometry," *Eur. Respiratory J.*, vol. 26, no. 2, pp. 319–38, 2005.

[3] C. P. Criée *et al.*, "Body plethysmography–Its principles and clinical use," *Respiratory Med.*, vol. 105, no. 7, pp. 959–971, Jul. 2006.

[4] R. Pierce, "Spirometry: An essential clinical measurement," *Australian Family Physician*, vol. 34, no. 7, pp. 535–539, Jul. 2005.

[5] V. Soleimani *et al.*, "Remote pulmonary function testing using a depth sensor," in *Proc. 2015 IEEE Biomed. Circuits Syst. Conf.*, Oct. 2015, pp. 1–4.

[6] V. Soleimani *et al.*, "Remote, depth-based lung function assessment," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 8, pp. 1943–1958, Aug. 2017.

[7] S. Ostadabbas *et al.*, "A passive quantitative measurement of airway resistance using depth data," in *Proc. 2014 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 5743–5747.

[8] S. Ostadabbas *et al.*, "A vision-based respiration monitoring system for passive airway resistance estimation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 9, pp. 1904–1913, Sep. 2016.

[9] H. Aoki *et al.*, "Non-contact respiration measurement using structured light 3-D sensor," in *Proc. 2012 SICE Annu. Conf.*, 2012, pp. 614–618.

[10] M.-C. Yu *et al.*, "Noncontact respiratory measurement of volume change using depth camera," in *Proc. 2012 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, pp. 2371–2374.

[11] T. M. Seppanen *et al.*, "Accurate measurement of respiratory airflow waveforms using depth data," in *Proc. 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 7857–7860.

[12] S. Rihana *et al.*, "Kinect2 – Respiratory movement detection study," in *Proc. 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 3875–3878.

[13] S. Transue *et al.*, "Real-time tidal volume estimation using iso-surface reconstruction," in *Proc. 2016 IEEE 1st Int. Conf. Connected Health, Appl., Syst. Eng. Technol.*, 2016, pp. 209–218.

[14] A. Chatterjee *et al.*, "Real-time respiration rate measurement from thoracoabdominal movement with a consumer grade camera," in *Proc. 2016 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2016, pp. 2708–2711.

[15] B. Reyes *et al.*, "Tidal volume and instantaneous respiration rate estimation using a smartphone camera," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 3, pp. 764–777, May 2017.

[16] C. Sharp *et al.*, "Remote pulmonary function testing–Computer gaming in the respiratory world," *Thorax*, vol. 70, pp. 117–118, 2015.

[17] C. Sharp *et al.*, "Towards respiratory assessment using depth measurements from a time-of-flight sensor," *Front. Physiol.*, vol. 8, 2017, Art. no. 65.

[18] Z. Zhang, "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 8, pp. 1902–1910, Aug. 2015.

[19] Z. Zhang *et al.*, "Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 522–531, Feb. 2015.

[20] G. Cennini *et al.*, "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction," *Opt. Express*, vol. 18, no. 5, pp. 4867–4875, 2010.

[21] Y. Sun *et al.*, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise," *J. Biomed. Opt.*, vol. 16, no. 7, 2011, Art. no. 077010.

[22] R. Yousefi *et al.*, "A motion-tolerant adaptive algorithm for wearable photoplethysmographic biosensors," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 2, pp. 670–681, Mar. 2014.

[23] E. Khan *et al.*, "A robust heart rate monitoring scheme using photoplethysmographic signals corrupted by intense motion artifacts," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 550–562, Mar. 2016.

[24] D. Shao *et al.*, "Noncontact monitoring breathing pattern, exhalation flow rate and pulse transit time," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 11, pp. 2760–2767, Nov. 2014.

[25] S. Beck and B. Froehlich, "Volumetric calibration and registration of multiple RGBD-sensors into a joint coordinate system," in *Proc. IEEE Symp. 3D User Interfaces*, 2015, pp. 89–96.

[26] M. Kowalski *et al.*, "Livescan3d: A fast and inexpensive 3d data acquisition system for multiple kinect v2 sensors," in *Proc. Int. Conf. 3D Vision*, 2015, pp. 318–325.

[27] V. Soleimani *et al.*, "3d data acquisition and registration using two opposing kinects," in *Proc. Int. Conf. 3D Vision*, 2016, pp. 128–137.

[28] M. A. Cretikos *et al.*, "Respiratory rate: The neglected vital sign," *Med. J. Australia*, vol. 188, no. 11, pp. 657–659, 2008.

[29] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Math., Phys. Eng. Sci.*, vol. 454, pp. 903–995, 1998.

[30] J. M. Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[3] *Sensor Platform for HEalthcare in a Residential Environment.*